Int J Life Cycle Assess (2014) 19:405–416
DOI 10.1007/s11367-013-0641-x

POLICIES AND SUPPORT IN RELATION TO LCA

Stochastic multi-attribute analysis (SMAA) as an interpretation method for comparative life-cycle assessment (LCA)

Valentina Prado-Lopez · Thomas P. Seager · Mikhail Chester · Lise Laurin · Melissa Bernardo · Steven Tylock

Received: 4 February 2013 / Accepted: 5 August 2013 / Published online: 3 September 2013

Abstract

Purpose Comparative life-cycle assessments (LCAs) today lack robust methods of interpretation that help decision makers understand and identify tradeoffs in the selection process. Truncating the analysis at characterization is misleading and existing practices for normalization and weighting may unwittingly oversimplify important aspects of a comparison. This paper introduces a novel approach based on a multi-criteria decision analytic method known as stochastic multi-attribute analysis for life-cycle impact assessment (SMAA-LCIA) that uses internal normalization by means of outranking and exploration of feasible weight spaces.

Methods To contrast different valuation methods, this study performs a comparative LCA of liquid and powder laundry detergents using three approaches to normalization and weighting: (1) characterization with internal normalization and equal weighting, (2) typical valuation consisting of external normalization and weights, and (3) SMAA-LCIA using outranking normalization and stochastic weighting. Characterized results are often represented by LCA software with respect to their relative impacts normalized to 100 %. Typical valuation approaches rely on normalization references, single value weights, and utilizes discrete numbers throughout the calculation process to generate single scores. Alternatively, SMAA-LCIA is capable of exploring high uncertainty in the input parameters, normalizes internally by pair-wise comparisons (outranking) and allows for the stochastic exploration of weights. SMAA-LCIA yields probabilistic, rather than discrete comparisons that reflect uncertainty in the relative performance of alternatives.

Results and discussion All methods favored liquid over powder detergent. However, each method results in different conclusions regarding the environmental tradeoffs. Graphical outputs at characterization of comparative assessments portray results in a way that is insensitive to magnitude and thus can be easily misinterpreted. Typical valuation generates results that are oversimplified and unintentionally biased towards a few impact categories due to the use of normalization references. Alternatively, SMAA-LCIA avoids the bias introduced by external normalization references, includes uncertainty in the performance of alternatives and weights, and focuses the analysis on identifying the mutual differences most important to the eventual rank ordering.

Conclusions SMAA-LCIA is particularly appropriate for comparative LCAs because it evaluates mutual differences and weights stochastically. This allows for tradeoff identification and the ability to sample multiple perspectives simultaneously. SMAA-LCIA is a robust tool that can improve understanding of comparative LCA by decision or policy makers.

Keywords Comparative life-cycle assessment · Decision analysis · Normalization · Outranking · Valuation

Responsible editor: Adriana Del Borghi

Electronic supplementary material The online version of this article (doi:10.1007/s11367-013-0641-x) contains supplementary material, which is available to authorized users.

V. Prado-Lopez (✉) · T. P. Seager · M. Chester · M. Bernardo
School of Sustainable Engineering and the Built Environment, Arizona State University, P.O. Box 9309, Tempe, AZ 85287-9309, USA
e-mail: valentina.prado@asu.edu

L. Laurin
EarthShift LLC, 31 Leach Road, Kittery, ME 03904, USA

S. Tylock
Sustainable Intelligence, c/o Adair Law Firm, 290 Linden Oaks, Suite 220, Rochester, NY 14625, USA

1 Introduction

Methodological challenges in normalization and weighting have received comparatively less research attention than those of

inventory building and characterization in life-cycle assessment (LCA). According to the International Standardization Organization (ISO), normalization and weighting are optional steps that require justification from LCA practitioners (ISO 14044 2006). Although the ISO guidelines mention normalization by means of a reference (external normalization) or by a baseline (internal normalization), in practice, the LCA community applies external normalization. Therefore, typical valuation as defined by this study is representative of methods such as ReCiPe, IMPACT 2002+, TRACI, and Ecoindicator—all which normalize externally (Lautier et al. 2010).

However, it is now recognized that problems in existing external normalization approaches include reference data gaps (Heijungs et al. 2007), a lack of consensus in data compilation (Bare et al. 2006), lack of uncertainty information (Lautier et al. 2010), and spatial and temporal variability (Finnveden et al. 2009; Bare and Gloria 2006). In addition, normalization references can be outdated, partly because compilation is a resource-intensive process. For example, the latest USA normalization reference, Tool for the Reduction and Assessment of Chemical and other Environmental Impacts, released in 2006 has a reference year of 1999 (Bare et al. 2006).

In response to these shortcomings, current research efforts in normalization focus on repairing and building normalization references and in creating approaches to document spatial and temporal discrepancies (Lautier et al. 2010; White and Carty 2010). Nonetheless, even if current issues with normalization reference datasets are resolved, typical valuation approaches with regard to normalization and weighting remain mathematically incompatible for comparative LCAs, where the goal is to identify an environmental preferable product, process, or pathway from a set of comparable alternatives with the same functional unit (Prado et al. 2012). In fact, the use of external normalization references in a comparative LCA can mask important aspects of a decision problem because the normalized impact depends on the size of the normalization reference (White and Carty 2010). This effect is evident when the normalization step completely overcomes the weights elicited from stakeholders or decision makers (Rogers and Seager 2009). For example, when using a normalization reference that includes a large inventory of emissions in a specific category, external normalization relative to that reference will systematically diminish differences between alternatives that might nevertheless be important to decision makers. To some, masking the environmental consequences of their choices by dividing them by emissions attributable to others may be the moral equivalent of justifying bad behavior by saying, “But everybody is doing it!”

Subjectivity concerns in the normalization and weighting stages of impact assessment often lead LCA practitioners to truncate impact assessment at characterization. While this may be effective for LCA motivated by improvement assessment, in a comparative LCA the characterized data present decision makers with too much information to interpret (Le Téno 1999; Boufateh et al. 2011). As a result, decision makers are forced to confront uncertain multi-criteria environmental problems without the aid of analytic guideposts, and may be subject to systematic biases, vulnerable to first impressions or prior stigmatization (Hertwich and Hammit 2001). To work around these difficulties, LCA practitioners may use the comparative impact representations built into several popular LCA software applications. These show the relative performance of a characterized inventory for each alternative, normalized so that 100 % in any impact category represents the worst performer among all the alternatives. In contrast to the ISO recommendations, this *internal normalization* approach avoids the necessity of external normalization references. However, these approaches lead to an analysis that is insensitive to magnitude, incapable of identifying tradeoffs (Norris 2001), and incorrectly presented as “unweighted” when in fact they represent *equal* weights that may or may not correspond to decision maker priorities. Thus, relying on default graphical outputs can be misleading.

There is an acute need for normalization and weighting methods in life-cycle impact assessment (LCIA) that can guide a comparative decision-making process in a transparent manner where subjective choices are not masked by modeling biases. The incorporation of a decision analysis framework to the interpretation stages is not intended to make decisions without the input from decision makers but rather help decision makers identify the areas of major tradeoffs among choices. This paper introduces a novel approach to normalization and weighting based on stochastic multi-attribute analysis (SMAA) that uses internal normalization by means of outranking and stochastic exploration of weight sets that do not privilege one impact category over others, or favor single viewpoints (Tylock et al. 2012). The method elucidates the trade-offs inherent in a comparative LCA problem, does not rely on external databases, and facilitates a more thorough exploration of uncertainty (including uncertainty and variability in weights among multiple stakeholders or decision makers). To illustrate application of the new method to a problem in comparative LCA, we present a case study in dry versus concentrated liquid laundry detergents using both typical and the novel approach to valuation.

1.1 LCA and decision analysis

Although it is widely understood that problems in comparative LCA present as paradigmatic multi-criteria-decision analytic problems under uncertainty, common valuation practices fail to incorporate knowledge from the fields of operations research, or decision analysis that might be brought to bear in LCA. Partly, this may be due to unresolved controversies within the decision analytic community itself. There are currently two schools of thought: normative and descriptive. The normative suggests that decision analysis should conform to

idealized mathematical or economic representations of how decisions *should* be made, whereas the descriptive maintains that decision analytics should be representative of the more heuristic and naturalistic processes that people *actually use* when confronting problems unaided. Both approaches are not intended to make decisions, rather *guide* the decision-making process in an iterative manner. However, it is vital to choose an approach that best fits the context of the problem. External normalization, as mentioned by ISO, more closely aligns with the *normative* school, while internal normalization also in ISO and present in SMAA more closely aligns with the *descriptive* school. Each approach has different assumptions and implications (Prado et al. 2012), as summarized in Table 1.

Outranking algorithms (and consequently SMAA) use pair-wise comparisons to assess the significance of mutual differences. The comparative performance of multiple alternatives are evaluated against pseudo-criteria called preference (p) and indifference (q) thresholds (Brans and Mareschal 2005), respectively representing the smallest difference between the performance of two alternatives on a single criterion that results in a conclusive preference for one over the other, and the largest difference that is entirely inconclusive (Rogers and Seager 2009). Thus, outranking allows the analyst to discard those categories in which the alternatives are deemed equivalent and focus attention on critical differences. Although internal normalization approaches in the absence of preference and indifference thresholds may result in “absurd” conclusions (Norris 2001), outranking avoids these pitfalls by distinguishing between negligible and significant differences (Prado et al. 2012). Moreover, because outranking relies on comparative pair-wise judgments, analysis can proceed with partially quantitative data, or even qualitative data (Gelderman and Schobel 2011).

SMAA combines outranking normalization with Monte Carlo analysis in weighting (Lahdelma and Salminen 2001; Lahdelma et al. 1998).

SMAA avoids subjectivity in weighting by allowing for the stochastic exploration of weight spaces rather than point values. Weighting in SMAA analysis concurs with the ISO definition of weights where “weighting is the process of converting indicator results of different impact categories by using numerical factors based on value-choices” (ISO 14044 2006). However, instead of applying discreet weight values based on averages or equal weights, this approach tests all possible perspectives without favoring a single value system (see stochastic weights in supplementary information). Currently, some LCA studies apply a “no weights” or “equal weights” approach where all impact categories are given the same weight value to avoid subjectivity biases and to avoid imposing a value system on others. However, this practice is in itself biased. There is a misconception that equal discreet weighting is a neutral position, when in fact this represents a very specific value system that may or may not reflect those values of the decision makers interpreting the results. Instead, a stochastic exploration of weights provides a useful analytic approach that is inclusive of all possible value systems and allows for the LCA analyst to view and present results without favoring a single value system. Stochastic weighting is not unique to SMAA, and can be applied separate of outranking normalization.

This study provides a more in depth analysis of the biases incurred by typical valuation practices and applies a more recent variation of SMAA (SMAA-TY), which constrains weight ranges with respect to the relative importance of each criterion for easier weight elicitation (Tylock et al. 2012). The advantage of SMAA-TY over SMAA is that the relative importance option in SMAA-TY weights allows decision makers to express the weights in terms of the level of priority, qualitatively, rather than quantifying it numerically. For example, decision makers can assign levels of priority to each impact category that range from “Well Below Average” to “Well Above Average.” This function allows for multiple parallel analyses in accordance to individual preferences of decision makers, where each decision maker generates a different set of results. This way, the analysis identifies the most and least preferred alternatives for each decision maker and this will allow decision makers to focus the discussion around the most conflicting alternatives to reach a balanced compromise. However, for the purposes of this study, all impact

Table 1 Normative and descriptive assumptions in external and outranking normalization

	External normalization in typical valuation	Outranking in SMAA
Underlying decision theory	Normative: models decision the way they *should* be made according to normative assumptions	Descriptive: models decision more like the way they *are* made
Type evaluation/measurement	Absolute	Relative (i.e., comparative)
Sensibility to context	Rigid. Evaluation of alternatives should not change if the options around it change (context independent)	Flexible. If alternatives change the relative evaluation of all alternatives may also change by causing a reframing of the problem (context dependent)
Appropriate application	Improvement assessment LCA	Comparative LCA

categories were categorized as "Average" since there were no defined preferences among impact categories.

1.2 Case study

The comparative LCA of laundry detergents used in this study covers the phases of raw material production, product manufacturing, packaging, transportation, and disposal of packaging. The use phase, retail, and product's end of life is equivalent in both formulations, thus are excluded from the LCA. The functional unit (FU) of the comparison is a standard dose of concentrated liquid and powder detergent. The detergents in this study represent typical market products of double concentration (2×) which can be found in retail stores in the USA.

The inventories of raw materials for the respective products use the ingredient formulations from the *Handbook of Detergents* (Showell 2006). The formulations in this study use point estimates of the content percent by weight of each chemical (see Electronic supplementary material (ESM)). Each chemical component is then matched to an appropriate Ecoinvent entry (dataset version 2.2) (Ecoinvent 2011). These entries contain chemical production requirements with respect to electricity, natural gas, and water (Koehler and Wildbolz 2009). For components not found in Ecoinvent, this model uses a proxy. For instance, enzymes in both formulations do not have an Ecoinvent dataset equivalent. Thus, this model uses average datasets of liquid enzyme with enzyme content of 4–6 % and a granular enzyme with an enzyme content of 4–6 % as proxies (Novozymes 2010). Inventory of packaging materials according to product surveys by The Sustainability Consortium (2011) is also sourced from Ecoinvent (see ESM).

Laundry detergent manufacturers double the concentration of their products because it allows for more doses in the same container, and improves storage, distribution, and transportation efficiency. Current consumer perceptions (as shaped by marketing messages) are that concentrated detergents save on packaging and transportation costs, and are therefore preferable from an environmental perspective. However, because laundry production is a wet chemical process, production of powder detergents requires an additional drying process prior to packaging—an energy intensive process that may nullify environmental gains in reduced packaging and transportation. Therefore, a comparative LCA can clarify whether the gains in transportation efficiency for powder detergents make up for the additional energy investment required in the manufacturing stage. After processing, the liquid alternative is packaged in a plastic bottle with a plastic cap and spout, and the powder alternative is packaged in a cardboard container with a plastic scoop (Table 2).

Both formulations are then distributed from the manufacturer to major cities across the USA. Transportation of detergents is based on an illustrative example in the USA. According to the US Census Bureau (2012), the largest economic activity in the laundry detergent industry occurs in Ohio. To explore the gains in transportation efficiency versus initial manufacturing energy investments, we model approximately the greatest transportation distance within contiguous USA—from Cincinnati, OH (headquarters of Procter and Gamble, a large laundry detergent manufacturer) to Los Angeles, CA (the most population dense city on the West coast). Thus, we assume the detergents travel approximately 2,300 mi (or 3,700 km) in heavy-duty trucks using diesel fuel. As both products are dense goods, fuel inventories are proportional to weight.

Based on the GREET 1.0 database, a heavy-duty truck has a load capacity of 25 short tons with a gas mileage of 5 mpg (Wang 2011). We assume each shipment to be at 90 % load capacity to take into account further tertiary packaging and pallets. The emissions during transportation depend on the fuel requirements per FU (Table 3). Further distribution to retail stores and use phase transportation is not included in the analysis.

Finally, the model includes disposal of the plastic and cardboard packaging of the detergents according to an average US waste scenario. Packaging is disposed of through a municipal solid waste system. Recycling rates according to the US Environmental Protection Agency (2011) are 19.3 % for low-density polyethylene bottles, 8.3 % for polypropylene

Table 2 Material and process inventory of each representative product for the liquid and powder detergent. The powder detergent contains more doses per packaged product

	Liquid laundry detergent	Powder laundry detergent
Main packaging—materials	170 g of high-density polyethylene bottle, polypropylene cap, and spout	355 g of cardboard box and polyethylene terephthalate scoop
Main packaging—processing	Bottle uses stretch blow molding and the cap and spout use an injection molding process	Cardboard box made from virgin material and cap uses injection molding
Number of FU per packaged product	64	80
Mass of functional unit	49.9 g of liquid detergent plus 2.6 g of combined packaging	34.1 g plus 4.4 grams of combined packaging

Therefore, the emissions from transportation per FU are about 40 % less for the powder detergent. GREET 1.0 emission factors/gal of diesel combusted can be found in the ESM

Table 3 Each heavy duty truck contains more doses of powder detergent than liquid

	Liquid detergent	Powder detergent
Weight per unit (includes product and packaging)	3.2 kg	2.7 kg
Number of units in one truck at 90 % capacity	7,080 bottles	8,400 boxes
Number of FU in one truck	453,120	672,000
Diesel fuel requirements for travel per FU	0.025 gal	0.017 gal

other packaging and 85 % of cardboard. There are no recycling credits given in this analysis because the impacts of the recycling process are attributed to the new product. However, this LCA does take into account the impacts of sending the remaining solid waste to landfills in accordance to the Ecoinvent dataset for sanitary landfill disposal. This model assumes no incineration. Finally, the impact of wastewater treatment for the residual laundry product is excluded as the active ingredients in each detergent are biochemically indistinguishable at the treatment plant.

To address uncertainty in the inventory data, this study uses the Pedigree Matrix. The Pedigree Matrix assumes a lognormal distribution (represented by arithmetic parameters) for each input in the model. The standard deviation of each distribution is based on six parameters: reliability, completeness, sample size and temporal, geographical and technological correlation (Weidema and Wesnæs 1996). Each parameter in the Pedigree Matrix can be described with a coefficient from 1 to 5. The matrix-based standard deviation captures uncertainties related to the assumptions of the input value. For instance, manufacturing data points tend to have less uncertainty because they are usually tightly controlled. While other inputs, such as transportation, have higher uncertainty because of their dependence on various factors like weather and traffic.

Therefore, this study assigns the pedigree coefficient to each input to model uncertainty. Ecoinvent data already contains the corresponding pedigree matrix coefficients that provide a standard deviation to the inventory. Then, a Monte Carlo simulation generates random values for each inventory input to populate the lognormal distribution. This simulation ran 350 scans for all inventories to ensure a complete depiction of the uncertainty values at a 90 % confidence level. The resulting inventories were characterized using the ReCiPe method, which multiplies the lognormal distribution of an inventory by a characterization factor represented by a single value. The ReCiPe impact assessment method characterizes the inventory into 18 midpoint impact categories (Table 4).

Table 4 Distributions of characterized results

Characterized Impact category	Unit	Liquid		Powder	
		Mean	SD	Mean	SD
Metal depletion	kg Fe eq	1.99E−04	2.28E−05	4.43E−03	4.64E−04
Water depletion	m^3	2.66E−03	4.42E−04	1.38E−03	1.74E−04
Terrestrial ecotoxicity	kg 1,4-DB eq	7.06E−04	1.81E−04	2.62E−04	6.97E−05
Agricultural land occupation	m^2	1.02E−02	2.84E−03	2.10E−02	3.80E−03
Ozone depletion	kg CFC-11 eq	9.24E−09	1.49E−09	6.65E−09	7.05E−10
Climate change	kg CO_2eq	9.04E−02	1.00E−02	1.02E−01	8.98E−03
Ionizing radiation	kg U_{235}eq	6.89E−03	8.91E−03	2.03E−02	1.28E−02
Marine ecotoxicity	kg 1,4-DB eq	5.10E−04	2.43E−04	1.04E−03	6.16E−04
Human toxicity	kg 1,4-DB eq	1.95E−02	1.21E−02	4.10E−02	2.41E−02
Urban land occupation	m^2	1.43E−03	4.75E−04	7.75E−04	6.39E−04
Freshwater ecotoxicity	kg 1,4-DB eq	7.02E−04	2.56E−04	1.17E−03	6.37E−04
Terrestrial acidification	kg SO_2eq	5.12E−04	1.95E−04	4.32E−04	6.54E−05
Marine eutrophication	kg N eq	7.05E−05	1.15E−05	7.93E−05	1.82E−05
Photochemical oxidant formation	kg NMVOC	3.06E−04	3.85E−05	2.89E−04	2.93E−05
Fossil depletion	kg oil eq	3.29E−02	7.44E−03	3.92E−02	1.91E−02
Freshwater eutrophication	kg P eq	3.49E−05	2.12E−05	4.20E−05	1.05E−04
Particulate matter formation	kg PM_{10}eq	1.55E−04	4.32E−05	1.57E−04	1.71E−05
Natural land transformation	m^2	1.95E−05	2.24E−04	1.15E−05	9.53E−05

The mean and standard deviations are both arithmetic, but the distribution is lognormal

2 Methods

To understand how SMAA-LCIA operates in comparison to other methods, this study compares three approaches to normalization and weighting in a comparative LCA of dry powder and concentrated liquid laundry detergents (Section 1.2). These three are: graphical outputs at characterization (resulting from internal normalization and equal weighting), typical valuation that consists of external normalization relative to national reference datasets, and SMAA-TY style valuation (Fig. 1).

2.1 Typical software output

Most comparative LCAs stop at the characterization stage to avoid subjectivity risks associated with valuation. However, avoiding valuation can lead to a misinterpretation of data. Uncertain characterized results are notoriously difficult to interpret unaided because of the large amount and disparate range of units and data ranges (as shown in Table 4). To speed interpretation of results, many studies represent characterized results in a single figure according to their relative performance in a 100 % scale with equal weights for all categories. This type of graph is also the main output from comparative analysis in most LCA software packages.

2.2 Typical valuation: external normalization and single value weights

In accordance with ISO recommendations, characterized inventory data can be normalized relative to an external normalization reference. These references are single values and report no uncertainty. Normalization is followed by weighting that evaluates multiple indicators according to the priorities of the decision makers' agendas. Weights are typically represented as single values without uncertainty and are entirely subjective because they depend on the individual priorities of decision makers (Schmidt and Sullivan 2002).

Overall, single scores follow Eq. (1) and utilize mean values of the characterized inventory. However, because categories are equally weighted in this study, the relative size of the weighted impacts is the same as the normalized impacts.

$$Single\ \ score = \sum \frac{Characterized\ \ Impact_i}{Normalization\ \ Reference_i} \times Weight_i \qquad (1)$$

2.3 Stochastic multi-attribute analysis TY

SMAA-TY, a modified version of SMAA, consists of stochastic outranking normalization (Fig. 2) and relative probabilistic weights (as shown in Tylock et al. 2012). Conventional SMAA methods examine the entire feasible weight space or a range (Lahdelma and Salminen 2001; Lahdelma et al. 1998), but SMAA-TY elicits weights in terms of the relative importance of each criterion according to six levels of importance: well above average, above average, average, below average, and well below average. This feature is unique to SMAA-TY. Instead of asking decision makers to translate preferences directly into numeric values or ranges, SMAA-TY facilitates weight elicitation because it requires qualitative inputs from decision makers. The choices by decision makers are then converted into a probability distribution as a function of the level of priority given, the total number of criteria and the confidence level. The confidence level in SMAA-TY is another unique feature and it ranges from Fair to Precise. The more confidence in a weight (i.e., the more precise), the more clustered the distribution. Likewise, if the confidence level is fair, the distribution is wider (programming information in the supporting information of Tylock et al. 2012). This study explores the entire weight space for all impact categories by giving all impact categories the same level of priority. This capability enables the analysis to test all possible perspectives

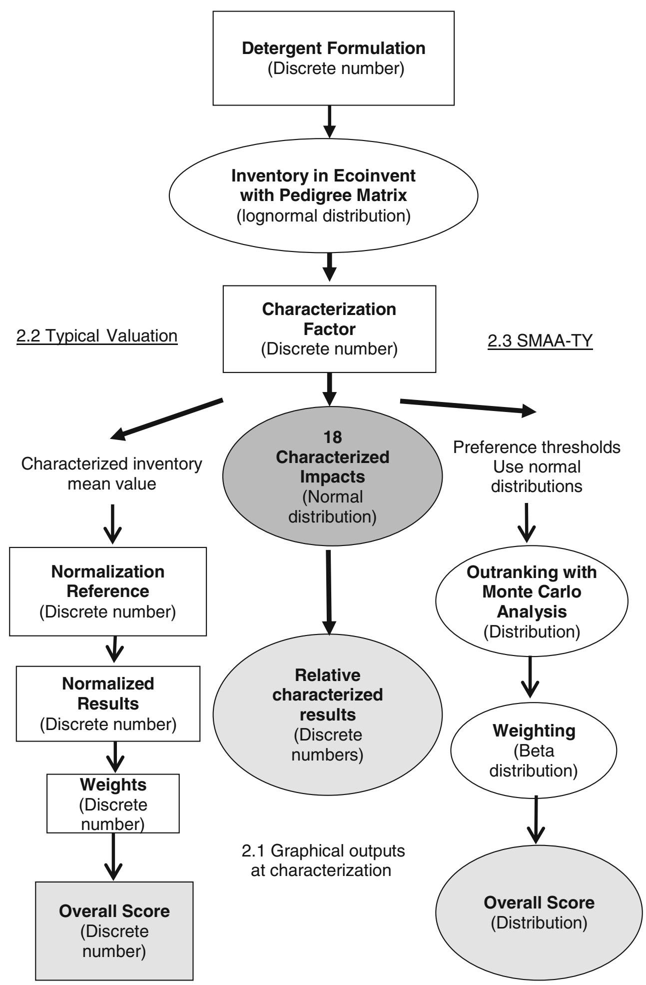

Fig 1 Three interpretation approaches; typical valuation, graphical output at characterization, and SMAA-TY valuation

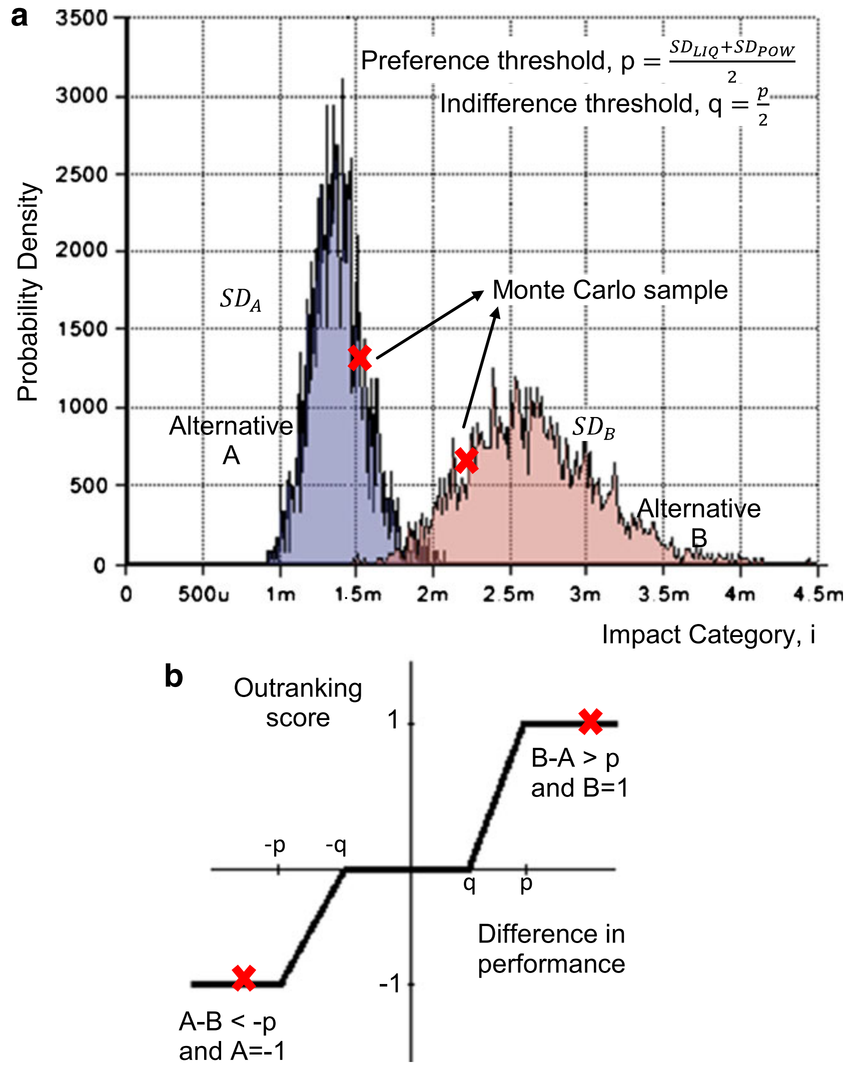

Fig 2 **a** *Top* shows the probabilistic performance of two hypothetical alternatives in impact category, i. Figure 5b (below), shows outranking function indicating the preference and indifference thresholds. Preference thresholds in this study equal the average between standard deviations, and the indifference threshold is half of the preference threshold. The difference in performance in each criterion in Fig. 5a is evaluated with the preference threshold (p) and the indifference threshold (q) in Fig. 5b. The outranking score ranges from −1 to 1 for each of the Monte Carlo runs (shown by the *red error mark*). This study performs 2,000 Monte Carlo simulations

without favoring single point estimates as in the case of discreet weight values or average estimations. In a separate context, there can be multiple parallel analyses in accordance to individual preferences of decision makers. These results would then make clear the tradeoffs for those with different values, providing a starting point for compromise.

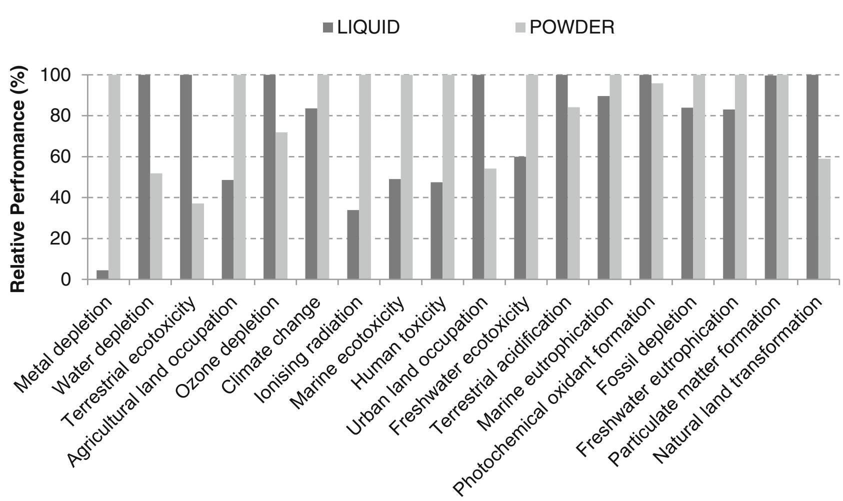

Fig 3 Eighteen characterized impact categories normalized to the worst in each category. Here, the liquid detergent performs relatively better in 11 of the 18 categories, whereas the powder alternative performs best in the remaining 7 impact categories. This graph lacks preference thresholds that measure the significance of “wins” and “losses”

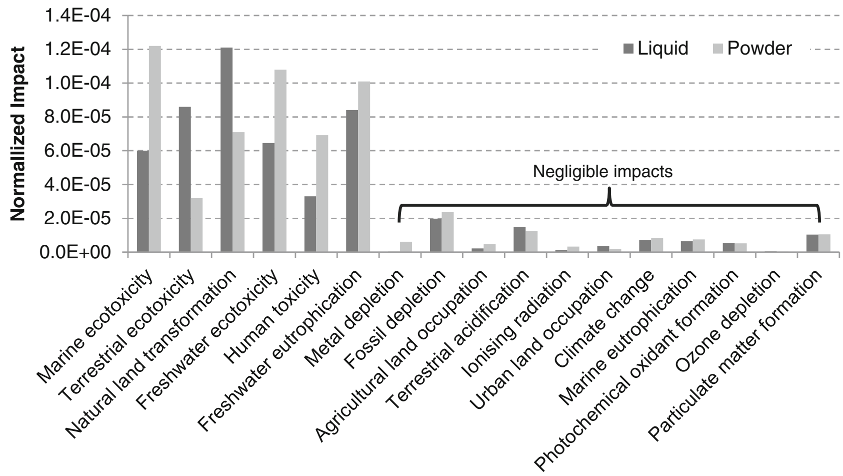

Fig 4 Normalized impacts according to the ReCiPe Midpoint Hierarchist model (values in supplementary information). No normalization reference that quantifies the regional water resources is available, thus it is not possible to include the water depletion category in the analysis of typical valuation (another drawback of relying in external databases). The impacts to the *left* of freshwater eutrophication have the greatest contributions, whereas the impacts to the *right* are negligible

Outranking scores are unitless numbers between −1 and + 1, where +1 is complete preference, and 0 is indifference (Behzadian et al. 2010; Figueira et al. 2005). The preference threshold (p) is the smallest difference between the two alternatives for which a complete preference may be inferred. In this case, a complete preference of +1 signifies an alternative that performs worse than the other in a given impact category. A score of −1 indicates superior comparative environmental performance. Indifference is determined by the *indifference threshold (q)*, the largest deviation considered negligible. Strict indifference occurs when both alternatives performances are within a negligible difference from each other (between $-q$ and$+q$), and both receive a 0. A weak preference is when the difference in performance lies between the indifference and preference threshold and it results in an interpolated value between −1 and +1. A weak preference means that an alternative is better than the other, but not enough to be a strict preference. Preference and Indifference thresholds can be selected through expert elicitation or with respect to the uncertainty of a given criteria (Linkov et al. 2007; Rogers and Bruen 1998; Rogers and Seager 2009). This study instead utilizes uncertainty in the data (shown in Table 4) to calculate preference and indifference thresholds according to Fig. 2a. Note that negative scores do not mean a negative impact (i.e., environmental benefit). Similar to typical valuation scores, a higher outranking score represents a greater environmental impact.

Before weighting, SMAA-TY winnows the alternatives to a maximum of eight impact categories in which the differences are the greatest and most significant. To evaluate this significance in the 18 characterized impact categories in Table 4, we derived the *relevance parameter* (r) from the outranking algorithm in Eq. (2). The relevance parameter is unique to this study, and it allows for a preliminary assessment

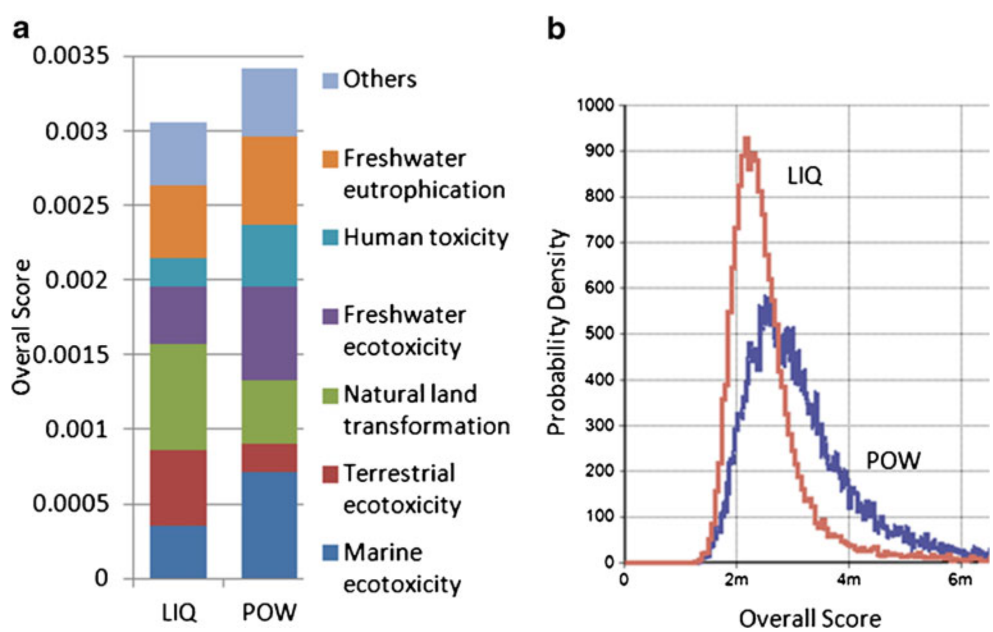

Fig 5 **a** *Left*, overall average and probabilistic scores for liquid (*LIQ*) and powder (*POW*) detergent according to the typical valuation method. The category "Others" is composed of metal depletion, fossil depletion, agricultural land occupation, terrestrial acidification, ionizing radiation, urban land occupation, climate change, marine eutrophication, photochemical oxidant formation, ozone depletion, and particulate matter formation. Probability distribution in Fig. (**b**, *right*) is a representation of the lognormal distribution of characterized impacts in Table 4

of the most influential categories at normalization via outranking.

$$r_i = \frac{\left|\mu_{LIQ} - \mu_{POW}\right|_i}{\frac{1}{2}\left(SD_{LIQ} + SD_{POW}\right)_i} \quad (2)$$

The relevance parameter uses data from the probability distribution of each impact assessment (see Table 4). The numerator in Eq. (2) represents the absolute difference between the means for each impact category, and the denominator represents the average of the arithmetic standard deviations for each impact category (i.e., the preference threshold as described in Fig. 2a). A large relevance parameter means that the alternatives mutual difference in performance is significant. It is important to include the standard deviations otherwise the absolute differences in mean values alone will favor those categories with greater magnitudes, when in fact their mutual difference might not be significant.

After identifying the most relevant categories, there are two steps in the SMAA-TY procedure. The first step involves performing 2,000 outranking Monte Carlo simulations for each of the eight impact categories and gathering outranking scores in terms of a probability distribution. The second step involves multiplication of the probabilistic outranking scores for each impact category by the probabilistic weights selected in the simulation. All weights were set at the same level of importance of "Average" with a confidence level of "Fair" (see ESM). Finally, SMAA-TY generates a probability distribution for the overall environmental score of each detergent.

3 Results

3.1 Typical software output

Figure 3 shows the relative performances of the liquid and powder detergent in 18 characterized impacts, where the better performing alternative is normalized relative to the poorer performer on a category-by-category basis. Figure 3 is insensitive to the magnitude (or significance) of each impact category, graphically depicting the worst performer in all categories as 100 % regardless of the absolute scale of emissions in any category. This graph can easily be misinterpreted because it suggests an approach of counting winners and losers in each category without any notion regarding the significance of those wins or losses. Without preference or indifference thresholds, this graph is unable to distinguish those differences that are important to decision makers from those that have such a small magnitude that they may reliably be ignored.

Even though this analysis is performed when *avoiding* valuation, there is an inherent valuation in Fig. 3 that analysts often fail to make explicit. While it is sometimes reported that output such as Fig. 3 is "unweighted," in fact the graphical depiction represents *equal* weights applied to each impact category (which is itself a subjective judgment). As a result, there is a need for valuation methods that process data in a way that highlights salient aspects without introducing subjectivity.

3.2 Typical valuation: external normalization and single value weights

Figure 4 reports the results of applying normalization factors inherent in the ReCiPe Midpoint Hierarchist methodology on a European scale within the SimaPro software package for the liquid and powder detergent. We apply equal weights for the

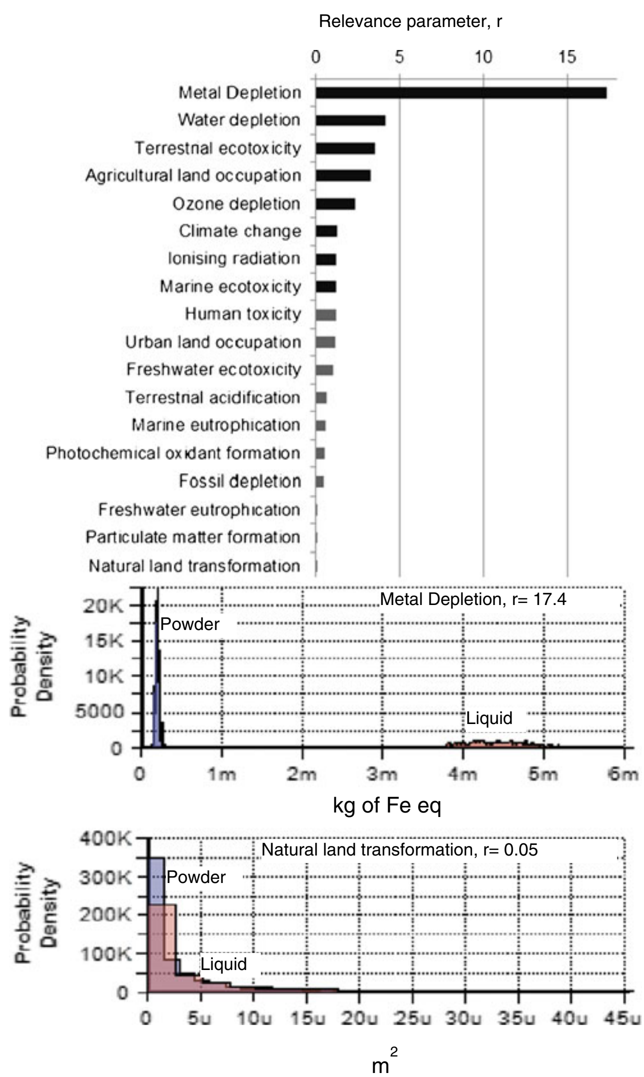

Fig 6 The relevance parameter of impact categories and a graphical representation of the most and least relevant categories (metal depletion and natural land transformation, respectively) are shown. Therefore, when probability distributions overlap the difference in performances is irrelevant because there is not enough certainty that one outperforms the other

17 impact categories for which normalization references are available, weighting each at 5.88 % to give impacts the same level of priority (Fig. 5).

Figure 5a shows that in average, the liquid alternative likely has a lesser overall environmental impact. Furthermore, the combined mean scores of 11 impact categories (labeled as “Others”) has a joint contribution of approximately 15 %, and the remaining six impact categories drive majority of the results. Therefore, the use of normalization references is masking 11 out of the 17 impact categories. Figure 5b shows the lognormal distribution of scores. Even though mean values favor the liquid alternative, the probability distributions overlap, indicating instances in which the liquid detergent performs worse than the powder detergent. However, this overlap is not visible when reporting single scores.

3.3 Stochastic multi-attribute analysis TY

Figure 6 shows the relevance parameter of the 18 characterized impact categories as described by Eq. (2). As the relevance parameter is a function of the mutual differences, there is one indicator per impact category.

From this point forward, there is a clear difference between normalized results of typical valuation and SMAA-TY. Even though the mutual difference between the detergents is very significant for the metal depletion category (Fig. 6), its impact is masked when dividing it by an external normalization reference (see Fig. 4). Therefore, evaluating performance with respect to a normalization reference fails to identify key tradeoffs among alternatives. In addition, the Water Depletion impact category which is not evaluated past characterization, in this case by ReCiPe, happens to be the second most relevant impact category (Fig. 6)—a significant impact that is excluded when there is a lack of a specific normalization reference.

Figure 7 shows the overall scores generated by SMAA-TY. The score reflects environmental impact, therefore the powder detergent, which is further to the right, is more likely to have a greater environmental impact. Scores are entirely relative to one another. Therefore, a negative score means a lower score with respect to other alternatives, not an environmental benefit.

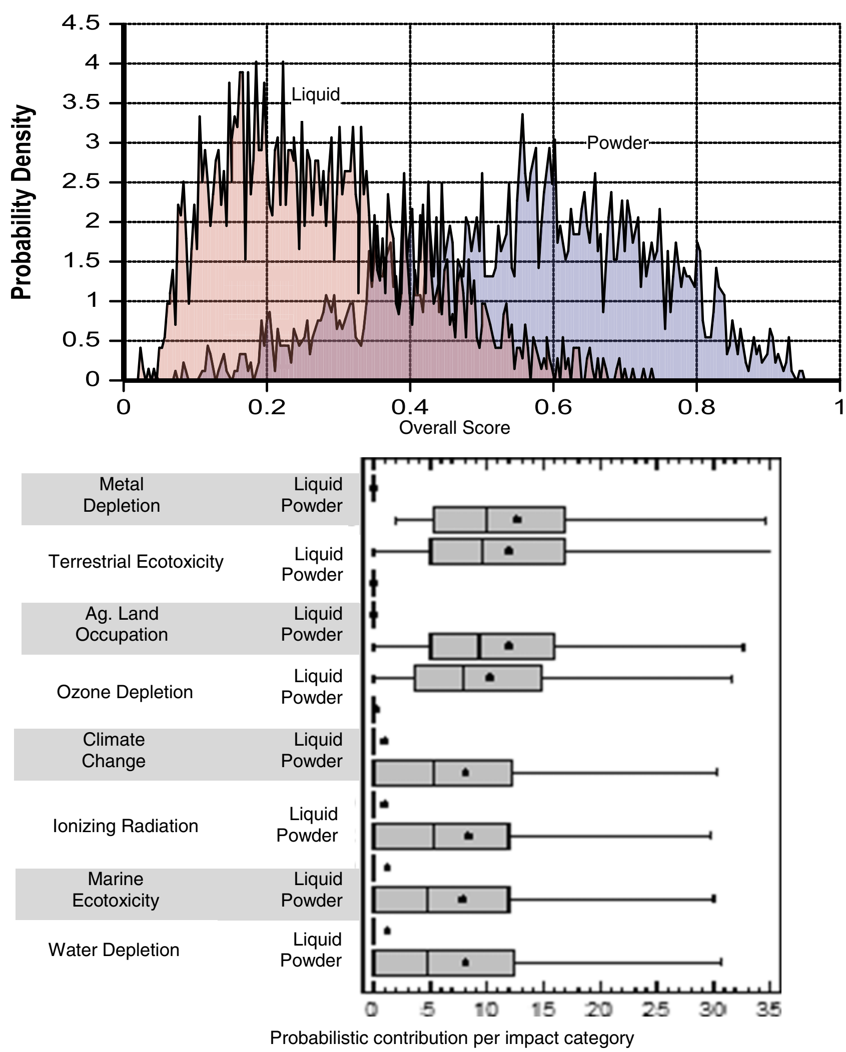

Fig 7 The probability distribution and contribution of the overall scores for the liquid and powder detergent according to SMAA-TY are shown

From this analysis, it can be calculated that the powder is 83 % likely to be worse than the liquid alternative. Alternatively, 17 % of the time the liquid detergent is worse than the powder alternative. The probabilistic score contributions in Fig. 7 show that five out of the eight most relevant impact categories evaluated in SMAA-TY can be masked by the use of normalization references. An additional category, Water Depletion, lacks normalization references in the ReCiPe method, so is not evaluated by the typical valuation approach.

4 Discussion

Both valuation methods recommend the liquid over the powder detergent, suggesting that the additional energy footprint during processing of the powder detergent exceeded the gains in transportation efficiency. However, the results shown in this paper are not intended to give a definite recommendation on laundry detergents. Rather, they highlight the capabilities of different interpretation approaches.

The first approach consists of showing characterized impacts in relation to one another (see Fig. 3). These results have inherent normalization and weighting which, according to ISO guidelines, weighting should not be applied when sharing LCA results to the public. In addition, they are insensitive to magnitude and oblivious to negligible and significant differences between impact categories. Thus, this type of analysis has severe limitations in terms of decision support.

Results from typical valuation in Fig. 5a show the overall scores resulting from external normalization on a European scale with single value egalitarian weighting. Results show that the categories accounting for at least 80 % of both total scores are: marine ecotoxicity, terrestrial ecotoxicity, natural land transformation, freshwater ecotoxicity, human toxicity, and freshwater eutrophication. The remaining *11* categories have little influence on both detergents. Single scores are a poor representation of life-cycle data because environmental performance is not a single indicator, nor is it a discrete value. Figure 5b shows the uncertainty in the overall scores of the typical valuation method. These scores remain distorted by normalization references.

Alternatively, the results from SMAA-TY show the probability distributions of the overall scores indicating the powder detergent is 83 % more likely to have a greater impact than the liquid detergent (see Fig. 7). The individual contribution from each weighted characterized impact category is also a distribution composed of 2,000 Monte Carlo runs represented by Box-and-Whisker plots. The breakdown by category in Fig. 7 shows that most impact categories contribute more to the powder detergent than to the liquid detergent. However, the liquid detergent has a greater impact in terrestrial ecotoxicity and ozone depletion. Five out of eight of the categories evaluated by SMAA-TY show a negligible contribution in the results from typical valuation in Fig. 4.

Out of the total 18 characterized impact categories generated by ReCiPe in Table 4, 6 impact categories drove the majority of the results in the typical valuation, and 8 impact categories drove the results in SMAA-TY. However, SMAA-TY focuses the analysis in the categories with the most tradeoffs—something that typical valuation and truncation at characterization fail to evaluate.

5 Conclusions

There is an acute need for interpretation methods in comparative LCAs that help understand the significance of mutual performances even before weighting. Graphical outputs at characterization fail at evaluating tradeoffs at the characterization stage, and typical practices in valuation, although in accordance to ISO guidelines, further distort data at normalization and weighting. This method conceals most impact categories due to the use of external normalization references. Therefore, it allows for a small fraction of the categories to dominate both scores and it does not present a robust platform for decision makers. Such high bias overcomes earlier efforts of data collection, impact assessment, and inputs from decision makers. An ideal valuation method is capable of guiding the decision-making process by revealing all aspects and dimensions of the problem in a transparent and concise way. A descriptive approach to interpretation that implements preference thresholds distinguishes between negligible and significant differences and can better highlight existing tradeoffs. We propose utilizing SMAA-TY based valuation applicable in all comparative LCAs. This novel method avoids masking criteria, is independent of external databases, includes multiple perspectives, and generates results that provides better information to decision makers.

Acknowledgments A previous version of this paper was presented at the 2013 International Symposium on Sustainable Systems and Technologies. This draft has benefited from the constructive comments made by the audience. In addition, the authors would like to thank the Sustainable Energy & Environmental Decision Science studio at Arizona State University for support throughout the preparation of this manuscript.

References

Bare J, Gloria T (2006) Critical analysis of the mathematical relationships and comprehensiveness of life cycle impact assessment approaches. Environ Sci Technol 40(4):1104–1113

Bare J, Gloria T, Norris G (2006) Development of the method and us normalization database for life cycle impact assessment and sustainability metrics. Environ Sci Technol 40:5108–5115

Behzadian M, Kazemzadeh RB, Albadvi A, Aghdasi M (2010) PROMETHEE: a comprehensive literature review. Eur J Oper Res 200:198–215

Boufateh I, Perwuelz A, Rabenasolo B (2011) Multiple criteria decision-making for environmental impacts optimization. Int J of Business Performance and Supply Chain Modeling 3(1): 28–42

Brans JP, Mareschal B (2005) PROMETHEE methods. Multiple criteria decision analysis: state of the art surveys. Springer, New York, pp 163–186

Census Bureau (2012) Industry Statistics NAICS 325611, U.S Department of Commerce. http://www.census.gov

Ecoinvent v2.2 (2011) Swiss Centre for Life Cycle Inventories. www.ecoinvent.org

Figueira J, Mousseau V, Roy B (2005) ELECTRE methods. Int Ser Oper Res Man 78(3):133–162

Finnveden G, Hauschild MZ, Ekvall T et al (2009) Recent developments in life cycle assessment. J Environ Manage 9:1–21

Gelderman J, Schobel A (2011) On the similarities of some multicriteria decision analysis methods. J MCDA 18(3–4):219–230

Heijungs R, Guinee J, Kleijn R et al (2007) Bias in normalization: causes, consequences, detection and remedies. Int J Life Cycle Assess 12(4):211–216

Hertwich EG, Hammit JK (2001) A decision analysis framework for impact assessment, part I: LCA and decision analysis. Int J Life Cycle Assess 6(1):5–12

International Standardization Organization 14044 (2006) Environmental management—life cycle assessment—requirements and guidelines.

Koehler A, Wildbolz C (2009) Comparing the environmental footprints of home-care and personal-hygiene products: the relevance of different life-cycle phases. Environ Sci Technol 43(22):8643–8651

Lahdelma RU, Salminen P (2001) SMAA-2: stochastic multicriteria acceptability analysis for group decision making. Oper Res 49(3):444–454

Lahdelma RU, Hokkanen J, Salminen P (1998) SMAA—stochastic multiobjective acceptability analysis. Eur J Oper Res 106:137–143

Lautier A, Rosenbaum RK, Margni M et al (2010) Development of normalization factors for Canada and the United States and comparison with European factors. Sci Total Environ 409(1):33–42

Le Téno JF (1999) Visual data analysis and decision support methods for nondeterministic LCA. Int J of Life Cycle Assess 4(1):41–47

Linkov I, Satterstrom FK, Yatsalo B et al (2007) Comparative assessment of several multi-criteria decision analysis tools for management of contaminated sediments. Environmental Security in Harbors and Coastal Areas. Springer, Amsterdam, pp 195–215

Norris GA (2001) The requirement for congruence in normalization. Int J Life Cycle Assess 6(2):85–88

Novozymes (2010) Liquid Enzyme Product LCI Dataset. Granulated Enzyme Product LCI Dataset. Novozymes

Prado V, Rogers K, Seager TP (2012) Integration of MCDA tools in valuation of comparative life cycle assessment. In: Curran MA (ed) Life cycle assessment handbook: a guide for environmentally sustainable products. John Wiley & Sons, Inc., Hoboken, NJ, USA. doi:10.1002/9781118528372.ch19

Rogers M, Bruen M (1998) Choosing realistic values of indifference, preference and veto thresholds for use with environmental criteria within ELECTRE. Eur J Oper Res 107:542–551

Rogers K, Seager TP (2009) Environmental decision-making using life cycle impact assessment and stochastic multi-attribute decision analysis: a case study on alternative transportation fuels. Environ Sci Technol 43(6):1718–1723

Schmidt WP, Sullivan J (2002) Weighting in life cycle assessments in a global context. Int J Life Cycle Assess 7(1):5–10

Showell EMS (2006) Handbook of detergents, part D: formulation. CRC Press, Boca Raton

The Sustainability Consortium (2011) Product Category Life Cycle Assessment (PCLCA) Laundry Detergent. The Sustainability Consortium "Sustainability Measurement and Reporting System" pilot project

Tylock SM, Seager TP, Snell J et al (2012) Energy management under policy and technology uncertainty. Energ Policy 47:156–163

US Environmental Protection Agency (2011) Municipal solid waste generation, recycling, and disposal in the United States—Table and Figures for 2010

Wang M (2011) GREET 1.0 Software. Center for Transportation Research, Energy Systems Division, Argonne National Laboratory. Copyright © 1999 U Chicago Argonne, LLC

Weidema BP, Wesnæs MS (1996) Data quality management for life cycle inventories — an example of using data quality indicators. J Clean Prod 4:167–174

White P, Carty M (2010) Reducing bias through process inventory dataset normalization. Int J Life Cycle Assess 15:994–1013